Multimodal diffusion geometry by joint diagonalization of Laplacians

Davide Eynard
Institute of Computational Science
Faculty of Informatics
Università della Svizzera italiana
Via G. Buffi 13, Lugano 6904, Switzerland
davide.eynard@usi.ch

Klaus Glashoff
Department of Mathematics
University of Hamburg
Hamburg, Germany
klaus.glashoff@math.uni-hamburg.de

Michael M. Bronstein
Institute of Computational Science
Faculty of Informatics
Università della Svizzera italiana
Via G. Buffi 13, Lugano 6904, Switzerland
michael.bronstein@usi.ch

Alexander M. Bronstein
School of Electrical Engineering
Faculty of Engineering
Tel Aviv University
Ramat Aviv 69978, Israel
bron@eng.tau.ac.il

Abstract

We construct an extension of diffusion geometry to multiple modalities through joint approximate diagonalization of Laplacian matrices. This naturally extends classical data analysis tools based on spectral geometry, such as diffusion maps and spectral clustering. We provide several synthetic and real examples of manifold learning, retrieval, and clustering demonstrating that the joint diffusion geometry frequently better captures the inherent structure of multi-modal data. We also show that many previous attempts to construct multimodal spectral clustering can be seen as particular cases of joint approximate diagonalization of the Laplacians.

1 Introduction

The Laplacian operator and related constructions play a pivotal role in a wide range of applications in machine learning, pattern recognition, and computer vision community. It has been shown that many problems in these fields boil down to finding some eigenvectors and eigenvalues of a Laplacian constructed on some high-dimensional data. Important examples include *spectral clustering* (Ng et al. (2001)) where clusters are determined by the first eigenvectors of the Laplacian; *eigenmaps* (Belkin & Niyogi (2002)) and more generally *diffusion maps* (Coifman & Lafon (2006)), where one tries to find a low-dimensional manifold structure using the first smallest eigenvectors of the Laplacian; and *diffusion metrics* (Coifman et al. (2005)) measuring the "connectivity" of points on a manifold and expressed through the eigenvalues and eigenvectors of the Laplacian. Other applications heavily relying on the properties of the Laplacian include *spectral graph partitioning* (Ding et al. (2001)), *spectral hashing* (Weiss et al. (2008)), spectral correspondence, image segmentation (Shi & Malik (1997)), and shape analysis (Levy (2006)). Because of the intimate relation between the Laplacian operator, Riemannian geometry, and diffusion processes, it is common to encounter the umbrella term *spectral* or *diffusion geometry* in relation to the above problems.

These applications have been considered mostly in the context of uni-modal data, i.e., a single data space. However, many applications involve observations and measurements of data done using different modalities, such as multimedia documents (Weston et al. (2010); Rasiwasia et al. (2010); McFee & Lanckriet (2011)), audio and video (Kidron et al. (2005); Alameda-Pineda et al. (2011)),

or medical imaging modalities like PET and CT (Bronstein et al. (2010)). Such problems of multimodal (or multi-view) data analysis have gained increasing interest in the computer vision and pattern recognition communities, however, there have been only few attempts extending the powerful spectral methods to such settings.

In this paper, we propose a general framework allowing to extend different diffusion and spectral methods to the multimodal setting by finding a common eigenbasis of multiple Laplacians. Numerically, this problem is posed as *approximate joint diagonalization* of several matrices. Such methods have received limited attention in the numerical mathematics community (Bunse-Gerstner et al. (1993)) and have been employed for joint diagonalization of covariance matrices in blind source separation applications by Cardoso & Souloumiac (1993, 1996); Yeredor (2002); Ziehe (2005). To the best of our knowledge, this is the first time they are applied to spectral embeddings. Besides providing a principled approach to data fusion, our framework gives a theoretical explanation to existing methods for multimodal data analysis. In particular, we show that many recent works on multi-view clustering by de Sa (2005); Ma & Lee (2008); Tang et al. (2009); Cai et al. (2011); Kumar et al. (2011) can be considered a particular instance of our framework.

2 Background

Let us be given some data represented as a k-dimensional manifold $X \subset \mathbb{R}^d$, embedded into a d-dimensional Euclidean space. In many applications d is very large while the intrinsic dimension of the data k is small, and one tries to study the structure of the manifold rather than its d-dimensional embedding. Such a structure can be characterized by the means of the *Laplace-Beltrami operator*. In the discrete setting, the manifold is often represented by a weighted graph with vertices $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset X$ and edge weights $w_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ representing local connectivity using e.g. Gaussian kernel (see von Luxburg (2007)). The Laplace-Beltrami operator can be discretized[1] as $\mathbf{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-1/2}$, where $\mathbf{W} = (w_{ij})$ and $\mathbf{D} = \mathrm{diag}(\sum_j w_{ij})$. Such a discretization is often referred to as *symmetric normalized Laplacian* and admits a unitary diagonalization $\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$, $\mathbf{V}\mathbf{V}^{\mathrm{T}} = \mathbf{I}_n$ with the eigenvalues $\lambda_1 = 0 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Geometric constructions associated with eigenvectors and eigenvalues of the Laplacian play an important role in machine learning, since several archetypical problems can be formulated in these terms:

Eigenmaps. Non-linear dimensionality reduction methods try to capture the intrinsic low-dimensional structure of the manifold X. Belkin and Niyogi (2002) showed that finding a neighborhood-preserving k-dimensional embedding of X can be posed as the minimum eigenvalue problem,

$$\min_{\mathbf{V} \in \mathbb{R}^{n \times k}} \mathrm{tr}\,(\mathbf{V}^{\mathrm{T}}\mathbf{L}\mathbf{V}) \;\; \text{s.t.} \;\; \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}. \tag{1}$$

This problem is minimized by setting $\mathbf{V}$ to be the matrix containing the first k eigenvectors of $\mathbf{L}$, thus effectively embedding the data by means of the eigenfunctions of the Laplace-Beltrami operator (the null eigenvector is usually discarded). Such an embedding is referred to as *Laplacian eigenmap*. More generally, a *diffusion map* is given as a mapping of the form $\Psi = (K(\lambda_2)\mathbf{v}_2, \ldots, K(\lambda_k)\mathbf{v}_k)$, where $K(\lambda)$ is some transfer function acting as a "low-pass filter" on eigenvalues λ (Coifman et al. (2005); Coifman & Lafon (2006)).

Diffusion distances. Coifman et al. (2005; 2006) related the eigenmaps to heat diffusion and random processes on manifolds and defined a family of *diffusion metrics* that in the most general setting can be written as

$$d^2(\mathbf{x}_i, \mathbf{x}_j) = \sum_l K(\lambda_l)(v_{il} - v_{jl})^2 = \|\Psi(\mathbf{x}_i) - \Psi(\mathbf{x}_j)\|_2^2. \tag{2}$$

Particular choice of $K(\lambda) = e^{-\lambda t}$ gives the *heat diffusion distance*, related to the connectivity of points $\mathbf{x}_i, \mathbf{x}_j$ on the manifold by means of diffusion process of length t. Such distances are intrinsic and thus invariant to manifold embedding and are robust to topological noise.

Spectral clustering. Ng et al. (2001) showed a very efficient and robust clustering approach based on the observation that the multiplicity of the null eigenvalue of $\mathbf{L}$ is equal to the number

[1] There exist many different constructions of the discrete Laplacian. For the sake of simplicity, we adopt the symmetric Laplacian. Our framework is applicable to other discretization as well.

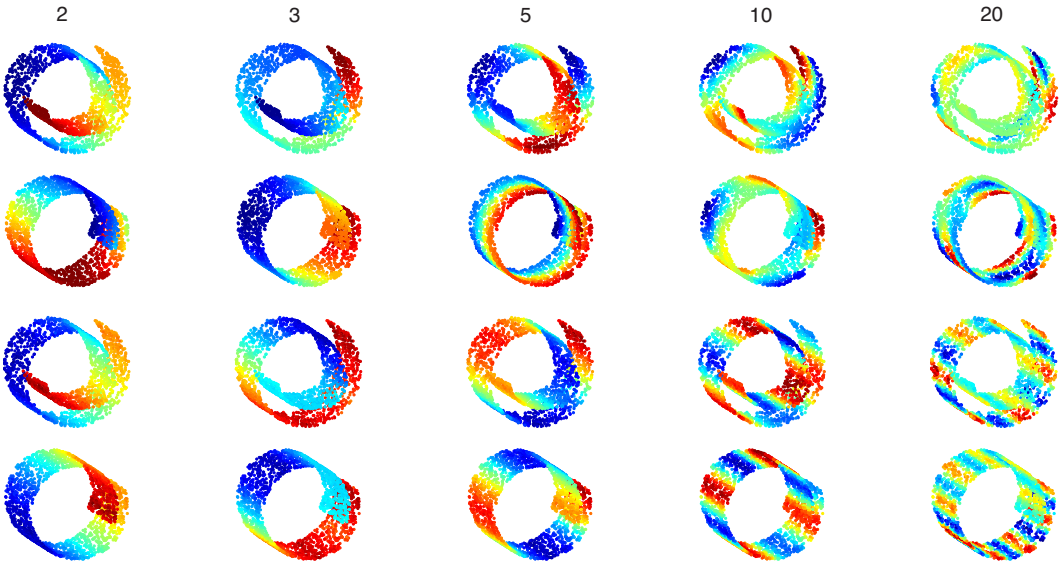


Figure 1: First and second rows: eigenfunctions of the Laplacians of two modalities of the Swiss roll. Third and fourth rows: joint eigenfunctions of the two Laplacians computed using JADE. Hot colors represent positive values; cold colors represent negative values.

of connected components of X. The corresponding eigenvectors act as indicator functions of these components. Embedding the data using these eigenvectors and then applying some standard clustering algorithm such as K-means was shown to produce significantly better results than clustering the high-dimensional data directly.

3 Multimodal diffusion geometry

Recently, we witness increasing popularity of attempts to analyze different "views" or modalities of data. Such data can be modeled as m different manifolds $X^1 \subset \mathbb{R}^{d_1}, \ldots, X^m \subset \mathbb{R}^{d_m}$, which can have embeddings of different dimensionality $(d_1, \ldots, d_m)$ and sometimes different structure. We are interested in analyzing these manifolds simultaneously in order to extract their joint intrinsic structure. We assume that we are given n corresponding samples $\{(\mathbf{x}_1^i, \ldots, \mathbf{x}_n^i) \subset \mathbb{R}^{d_i}\}_{i=1}^m$ on the manifolds and can construct the Laplacian matrices $\mathbf{L}_1, \ldots, \mathbf{L}_m$ as described in the previous section.

Trying to use the eigenvectors $\mathbf{V}_1, \ldots, \mathbf{V}_m$ of the Laplacian matrices $\mathbf{L}_1, \ldots, \mathbf{L}_m$ is problematic: for a set of eigenvectors corresponding to an eigenvalue with multiplicity greater than one, we can talk only of eigen sub-space, and any basis spanning it is a valid set of eigenvectors. As a result, the eigenvectors of the Laplacians in different modalities can be substantially different (Figure 1, top).

Joint diagonalization. A solution is to try to find the eigenbasis of the Laplacians *simultaneously*. This problem is known as *joint diagonalization* and consists of finding a set of joint orthogonal eigenvectors $\bar{\mathbf{V}}$ such that $\bar{\mathbf{V}}^{\mathrm{T}}\mathbf{L}_i\bar{\mathbf{V}} = \mathbf{\Lambda}_i$ are diagonal matrices of the eigenvalues of $\mathbf{L}_i$. Such a common eigenbasis solves the inherent ambiguity in the definition of the eigenvectors and "couples" different modalities (Figure 1, bottom). However, due to differences between the modalities and the presence of noise, the Laplacian matrices $\mathbf{L}_1, \ldots, \mathbf{L}_m$ rarely have a joint eigenbasis (iff they commute). It is still possible to find an *approximate* joint diagonalization by solving

$$\min_{\bar{\mathbf{V}}} \sum_{i=1}^{m} \mathrm{off}(\bar{\mathbf{V}}^{\mathrm{T}}\mathbf{L}_i\bar{\mathbf{V}}), \quad \text{s.t. } \bar{\mathbf{V}}^{\mathrm{T}}\bar{\mathbf{V}} = \mathbf{I}, \tag{3}$$

where $\mathrm{off}(\mathbf{X})$ is some off-diagonality criterion, e.g. the sum of squared off-diagonal elements, $\mathrm{off}(\mathbf{X}) = \|\mathbf{X} - \mathrm{diag}(\mathbf{X})\|_{\mathrm{F}}^2$. In this case, $\bar{\mathbf{V}}^{\mathrm{T}}\mathbf{L}_i\bar{\mathbf{V}}$ are only approximately diagonal; we refer to the average of the diagonal elements $\bar{\mathbf{\Lambda}} = \frac{1}{m}\sum_{i=1}^m \mathrm{diag}(\bar{\mathbf{V}}^{\mathrm{T}}\mathbf{L}_i\bar{\mathbf{V}})$ as the *joint approximate eigenvalues* of $\mathbf{L}_1, \ldots, \mathbf{L}_m$. This definition allows us to naturally extend the diffusion geometric

methods discussed in the previous section (eigenmaps, diffusion distances, spectral clustering, etc.) to the multimodal setting by simply replacing the eigenvalues and eigenvectors of a single Laplacian $\mathbf{L}_i$ by the joint eigenvectors $\bar{\mathbf{V}}$ and eigenvalues $\bar{\mathbf{\Lambda}}$ of multiple Laplacians $\mathbf{L}_1, \ldots, \mathbf{L}_m$.

Numerical computation. A numerical method for joint diagonalization based on a modified *Jacobi iteration* traces back to Bunse-Gerstner et al. (1993), and it has been used at about the same time by Cardoso and Souloumiac (1993; 1996) for joint diagonalization of covariance matrices in the context of blind source separation. The idea of the standard Jacobi method for eigenvalue calculation is to apply a sequence of plane rotations in order to sequentially minimize the off-diagonal elements of the given matrix. The rotation is applied "in-place" and does not require matrix multiplication. In the modified Jacobi method (referred to as JADE), the rotations are applied to reduce the off-diagonality criterion (3) in each step. Let $\mathbf{R}_{pqcs}$ the (complex) rotation matrix the entries of which are equal to those of the identity matrix except for the elements

$$\begin{pmatrix} r_{pp} & r_{pq} \\ r_{qp} & r_{qq} \end{pmatrix} = \begin{pmatrix} c & \bar{s} \\ -s & \bar{c} \end{pmatrix} \tag{4}$$

where $| c |^2 + | s |^2 = 1$. Cardoso & Souloumiac (1996) show that the problem

$$\min_{|c|^2+|s|^2=1} \sum_{i=1}^{m} \mathrm{off}(\mathbf{R}_{pqcs}^{\mathrm{T}} \mathbf{L}_i \mathbf{R}_{pqcs}) \tag{5}$$

has a simple explicit solution based on a 3×3 eigenvalue problem. JADE is one of the most common algorithms in the field of joint diagonalization and has complexity comparable to that of the standard Jacobi method. There are other algorithms, like the ACDC method of Yeredor (2002), as well as different versions of the idea of minimizing a suitable cost function on the Stiefel manifold (Rahbar & Reilly (2000)).

Analytic computation. In the spectral clustering problem, we are looking for the null eigenvectors of the Laplacian. Assuming that the first k eigenvalues of the Laplacians are zero, we want to find $\bar{\mathbf{V}} \in \mathbb{R}^{n \times k}$ such that $\mathbf{L}_i \bar{\mathbf{V}} = 0$ for all $i = 1, \ldots, m$ and $\bar{\mathbf{V}}^{\mathrm{T}} \bar{\mathbf{V}} = \mathbf{I}$ by reformulating (3) as

$$\min_{\bar{\mathbf{V}} \in \mathbb{R}^{n \times k}} \sum_{i=1}^{m} \| \mathbf{L}_i \bar{\mathbf{V}} \|_{\mathrm{F}}^2, \;\; \text{s.t.} \;\; \bar{\mathbf{V}}^{\mathrm{T}} \bar{\mathbf{V}} = \mathbf{I}. \tag{6}$$

Since $\sum_{i=1}^{m} \| \mathbf{L}_i \bar{\mathbf{V}} \|_{\mathrm{F}}^2 = \mathrm{tr}\,(\bar{\mathbf{V}}^{\mathrm{T}} (\sum_{i=1}^{m} \mathbf{L}_i^{\mathrm{T}} \mathbf{L}_i) \bar{\mathbf{V}})$, the problem can be equivalently recast as single-modality clustering with the "average" Laplacian matrix $\bar{\mathbf{L}} = \sum_{i=1}^{m} \mathbf{L}_i^{\mathrm{T}} \mathbf{L}_i$. We can also consider other averaging operators, e.g. weighted arithmetic mean $\bar{\mathbf{L}} = \sum_{i=1}^{m} w_i \mathbf{L}_i$ or harmonic mean $\bar{\mathbf{L}} = (\sum_{i=1}^{m} \mathbf{L}_i^{-1})^{-1}$. We discuss these methods in the next section.

For zero eigenvalues, (6) is akin to (3), which justifies the successful use of such "averaging" methods in problems of multimodal spectral clustering (Ma & Lee (2008); Cai et al. (2011)). However, iterative methods such as JADE explicitly minimizing the off-diagonality criterion (3) are more generic and applicable to settings where one has to find all or many joint eigenvectors, e.g., for computing eigenmaps or diffusion distances.

4 Relation to previous works

There have been numerous recent works on multimodal spectral-type clustering proposing different ways of fusing multiple modalities based on different principles. Considering these methods through the prism of joint diagonalization, we show many commonalities and equivalences between algorithms stemming from different motivations and coming from various communities. Ma & Lee (2008) considered detection of shots in video sequences using fusion of video and audio information, employing for this purpose spectral clustering of a Laplacian created as a weighted arithmetic mean of each modality Laplacian. Tang et al. (2009) used low-rank factorization of the weight matrix, trying to find a common factor $\mathbf{U}$ such that $\mathbf{W}_i \approx \mathbf{U} \mathbf{\Lambda}_i \mathbf{U}^{\mathrm{T}}$ by solving

$$\min_{\mathbf{U} \in \mathbb{R}^{n \times k}, \Lambda_i \in \mathbb{R}^{n \times n}} \sum_{i=1}^{m} \| \mathbf{W}_i - \mathbf{U} \mathbf{\Lambda}_i \mathbf{U}^{\mathrm{T}} \|_{\mathrm{F}}^2, \tag{7}$$

using the quasi-Newton method. Besides the fact that the factorization is applied to the weight matrix (it can be equivalently applied to the Laplacian), we see here a (non-orthogonal) joint diagonalization problem with an off-diagonality criterion considered by Yeredor (2002).

Cai et al. (2011) proposed a method for *multiview spectral clustering* (MVSC) by solving[2]

$$\min_{\mathbf{V}_i,\mathbf{V}\in\mathbb{R}^{n\times k}} \sum_{i=1}^{m} \mathrm{tr}\left(\mathbf{V}_i^{\mathrm{T}}\mathbf{L}_i\mathbf{V}_i\right) + \alpha\|\mathbf{V}_i-\mathbf{V}\|_{\mathrm{F}}^2 \;\text{ s.t. }\; \mathbf{V}^{\mathrm{T}}\mathbf{V}=\mathbf{I} \tag{8}$$

The authors show that this problem can be equivalently posed as

$$\max_{\mathbf{V}\in\mathbb{R}^{n\times k}} \mathrm{tr}\left(\mathbf{V}^{\mathrm{T}}\textstyle\sum_{i=1}^{m}\left(\mathbf{L}_i+\alpha\mathbf{I}\right)^{-1}\mathbf{V}\right) \;\text{ s.t. }\; \mathbf{V}^{\mathrm{T}}\mathbf{V}=\mathbf{I}, \tag{9}$$

and then employ an iterative algorithm to find the solution $\mathbf{V}$. First, we observe that problem (8) consists of m minimum-eigenvalue problems w.r.t. bases $\mathbf{V}_i$, with the addition of a coupling term, encouraging $\mathbf{V}_i$ as close as possible to some common basis $\mathbf{V}$ (note that the authors do not impose orthogonality constraints $\mathbf{V}_i^{\mathrm{T}}\mathbf{V}=\mathbf{I}$, but for $\alpha\gg 0$, the proximity to orthogonal $\mathbf{V}$ makes $\mathbf{V}_i$ approximately orthogonal). Thus, it is possible to interpret (8) as a kind of joint diagonalization criterion. Second, problem (9) can be rewritten as a minimum eigenvalue problem

$$\min_{\mathbf{V}\in\mathbb{R}^{n\times k}} \mathrm{tr}\left(\mathbf{V}^{\mathrm{T}}\left(\textstyle\sum_{i=1}^{m}\left(\mathbf{L}_i+\alpha\mathbf{I}\right)^{-1}\right)^{-1}\mathbf{V}\right) \;\text{ s.t. }\; \mathbf{V}^{\mathrm{T}}\mathbf{V}=\mathbf{I}, \tag{10}$$

whose solution is given by the matrix composed of the first k eigenvectors of the matrix $\left(\sum_{i=1}^{m}\left(\mathbf{L}_i+\alpha\mathbf{I}\right)^{-1}\right)^{-1}$. For $\alpha>0$, this a regularized version of the *harmonic mean* of the Laplacian matrices. We can thus regard the method of Cai et al. (2011) as a particular instance of our joint diagonalization approach discussed in the previous section.

Kumar et al. (2011) proposed the *centroid co-regularization* approach for multimodal clustering based on the minimization of

$$\min_{\mathbf{V},\mathbf{V}_i\in\mathbb{R}^{n\times k}} \sum_{i=1}^{m} \mathrm{tr}\left(\mathbf{V}_i^{\mathrm{T}}\mathbf{L}_i\mathbf{V}_i\right) - \alpha\mathrm{tr}\left(\mathbf{V}_i\mathbf{V}_i^{\mathrm{T}}\mathbf{V}\mathbf{V}^{\mathrm{T}}\right) \;\text{ s.t. }\; \mathbf{V}_i^{\mathrm{T}}\mathbf{V}_i=\mathbf{I};\;\; \mathbf{V}^{\mathrm{T}}\mathbf{V}=\mathbf{I}. \tag{11}$$

This function is alternatingly minimized, first with respect to the $\mathbf{V}_i$, then with respect to $\mathbf{V}$. Problems (11) and (8) are similar in their spirit (the first one uses dissimilarity $\|\mathbf{V}_i-\mathbf{V}\|_{\mathrm{F}}^2$ as coupling term, while the second one the similarity $\mathrm{tr}\left(\mathbf{V}_i\mathbf{V}_i^{\mathrm{T}}\mathbf{V}\mathbf{V}^{\mathrm{T}}\right)=\|\mathbf{V}_i^{\mathrm{T}}\mathbf{V}\|_{\mathrm{F}}^2$)), and fall under our joint diagonalization framework.

We must stress that these methods were developed for clustering problems where one has to find the null eigenvectors, and do not adapt easily to other applications of diffusion geometry where one has to find many or all joint eigenvectors of the Laplacians (e.g., computation of diffusion distances). In particular, iterative solvers used in Tang et al. (2009); Kumar et al. (2011); Cai et al. (2011) do not scale up to such cases. On the other hand, algorithms such as modified Jacobi iteration (JADE) are made for finding a full set of joint eigenvectors and have the complexity akin to standard Jacobi iteration. Further speed-up might be achieved by making explicit use of the sparse structure of the Laplacian matrices, which is not taken advantage of in JADE.

5 Results

We tested the proposed approach on three applications: dimensionality reduction, diffusion distance, and spectral clustering. All the datasets and code generating the results in this section are available from `anonymous.com`. Additional results are shown in the supplementary material.

Swiss rolls. In the first experiment, we used two Swiss roll surfaces with slightly different embedding as two different data modalities. The rolls were constructed in such a way that in each modality

[2] Cai et al. (2011) also impose a non-negativity constraint on the matrix $\mathbf{V}$ in order to obtain cluster indicators directly and bypass the K-means clustering stage. We ignore this additional constraint for the simplicity of discussion; such a constraint can be added to all the problems discussed in this paper.

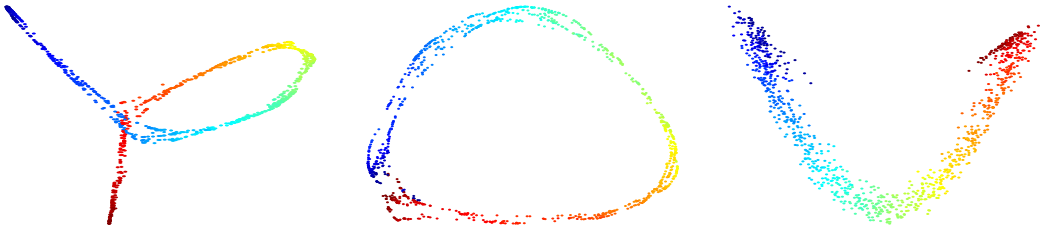
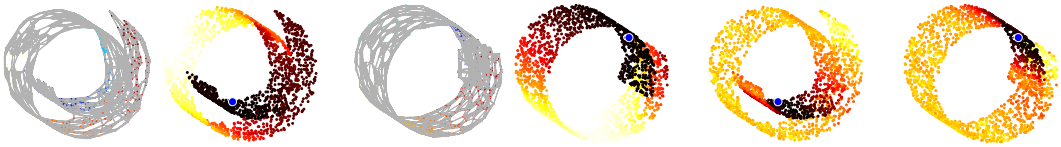


Figure 2: Flattening the Swiss rolls: dimensionality reduction using unimodal (left, center) and multimodal (right) eigenmaps. Joint eigenvectors were computed using JADE.



Modality 1 Modality 2 Multimodal

Figure 3: Diffusion distances from the blue point to the rest of the points on the Swiss roll surfaces. Darker colors represent smaller distances. First and third columns show the connectivity used in the construction of the Laplacians. Joint eigenvectors were computed using JADE.

there is topological noise (connectivity "across" the roll loops) at different points. Laplacians were constructed as in Belkin & Niyogi (2002) using 5-neighbor connectivity and Gaussian weights with scale parameter t. Figure 1 shows the first few eigenvectors computed using each Laplacian individually and jointly. Figure 2 shows two-dimensional embeddings of the same surfaces using the first non-trivial eigenvectors. When using joint eigenvectors, we are able to correctly capture the intrinsic structure of the data. Figure 3 shows the diffusion distance on the Swiss roll surfaces, computed using the first 100 eigenvectors and heat diffusion kernel $K(\lambda) = e^{-1000\lambda}$. Topological noise is clearly visible especially in the first modality, resulting in the distance between two loops to be small. This phenomenon does not occur when using joint eigenvectors.

Synthetic data clustering. In the second experiment, we performed clustering on several synthetic multimodal datasets. Laplacians were constructed using 15 nearest neighbors (10 for the *circles*), and Gaussian weight selected using the self-tuning approach of Perona & Zelnik-Manor (2004). We compare spectral clustering based on single modalities (SC-1 and SC-2) and joint diagonalization obtained using the JADE method of Cardoso & Souloumiac (1996); harmonic mean (JD-HM) of Laplacians (Cai et al. (2011)); and a non-spectral Comraf clustering algorithm (Bekkerman & Jeon (2007)). Quality was measured using the clustering accuracy criterion as defined in Bekkerman & Jeon (2007). For *Blobs*, accuracy is averaged over 100 experiments ran on randomly generated datasets.

The results are summarized in Figure 4 and Table 1. Surprisingly, the simple-minded averaging approach performs extremely well; this is consistent with the previously reported results and the success of the methods of Cai et al. (2011) (essentially harmonic mean) and Ma & Lee (2008) (arithmetic mean).

	Clus.	**SC-1**	**SC-2**	**JADE**	**JD-HM**	**Comraf**
Blobs	6	91.0±7.2%	90.8±7.2%	97.3±4.2%	98.3±3.0%	86.9±8.6%
Circles	4	65.9%	63.4%	100.0%	99.8%	31.4%
NIPS	4	63.3%	75.1%	99.9%	99.9%	51.8%
NUS	7	83.5%	71.0%	92.4%	80.7%	82.1%
Caltech	7	73.3%	76.2%	86.7%	84.8%	–
	20	66.3%	70.7%	73.3%	76.0%	–

Table 1: Accuracy of different clustering methods.

NUS dataset. In the third experiment, we used a subset of the NUS-WIDE dataset Chua et al. (2009) containing annotated images. The images were selected on purpose to have ambiguous con-

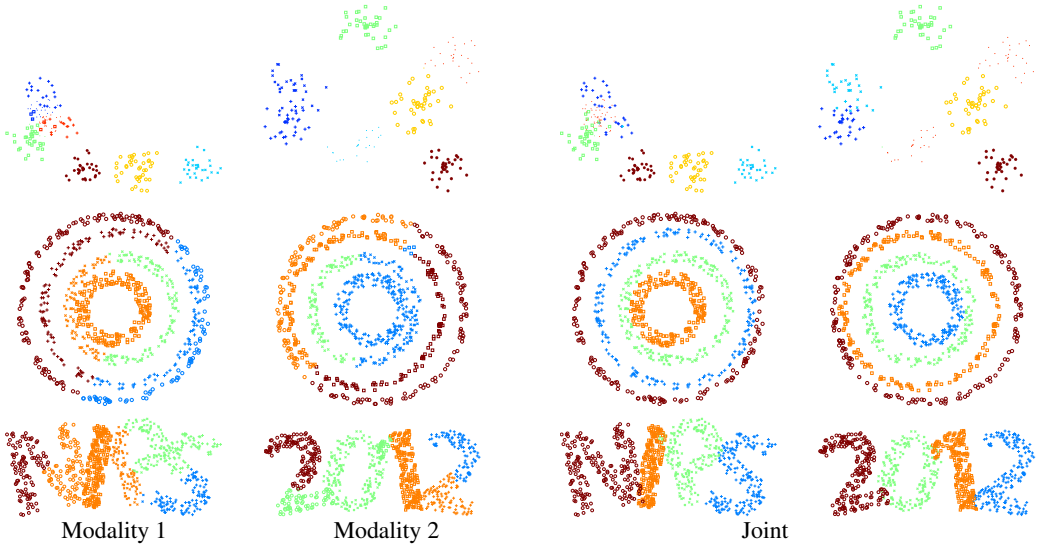


Figure 4: Clustering synthetics datasets. Marker size represents ground truth; marker color represents segmentation results (ideally, markers of each type should have a single color).

tent and annotations (e.g., swimming tigers are also tagged as "water" making them confuse e.g. with whales). As two different modalities, we used the 64-dimensional color histograms and 1000-dimensional bags of words. Laplacians were constructed using 10 nearest neighbors and Gaussian weight was selected using self-tuning. Table 1 shows the performance of different clustering methods, and Figure 5 exemplifies the clustered images.

Using JADE joint diagonalization, we produced all the joint eigenvectors of the two modalities Laplacians. Figure 8 (top) shows the distance matrices between the objects in the NUS dataset obtained using uni- and multi-modal diffusion distances (computed with the first 100 eigenvectors according to (2) using heat diffusion kernel $K(\lambda) = e^{-5\lambda}$). Ideally, the distance matrix should contain zero blocks on the diagonal (objects of the same class) and non-zero elsewhere (objects from different classes). Thresholding these distances at a set of levels and measuring the false positives/true positive rates (FPR/TPR), we produce the ROC curves that clearly indicate the advantage of using multiple modalities (Figure 8).

In Figure 7 (top), we used the diffusion distance to progressively sample the NUS dataset using the farthest point sampling strategy: starting with some point, pick up the second one as most distant from the first; then the third as the most distant from the first and second, and so on. Such sampling is almost-optimal (Hochbaum & Shmoys (1985)) and is known to produce a progressively refined r-covering of the set. In fact, the first 7 samples produced in this way cover all the classes present in the dataset, which is an indication of the meaningfulness of such a sampling.

Caltech dataset. In the fourth experiment, we repeated the third experiment on a subset of the Caltech-101 dataset with 7 and 20 image classes as in Cai et al. (2011). For each image, kernels arising from different visual descriptors were given. For the 7-clusters experiment, we used the bio-inspired features and 4x4 pyramid histogram of visual words (PHOW); for the 20-clusters experiment, we used geometric blur and 4x4 PHOW descriptors as different modalities, respectively. Laplacians were constructed from these kernels using Gaussian weight selected with self-tuning. Diffusion distances were computed with the first 100 eigenvectors using the kernel $K(\lambda) = e^{-5\lambda}$. The results are shown in Figures 6–8.

6 Discussion and Conclusions

We presented a framework for multi-modal data analysis using approximate joint diagonalization of Laplacian matrices, naturally extending the classical construction of diffusion geometry to the multi-modal setting. This construction allowed an almost straightforward extension of various diffusion-geometric data analysis tools such as spectral clustering and manifold learning based on diffusion

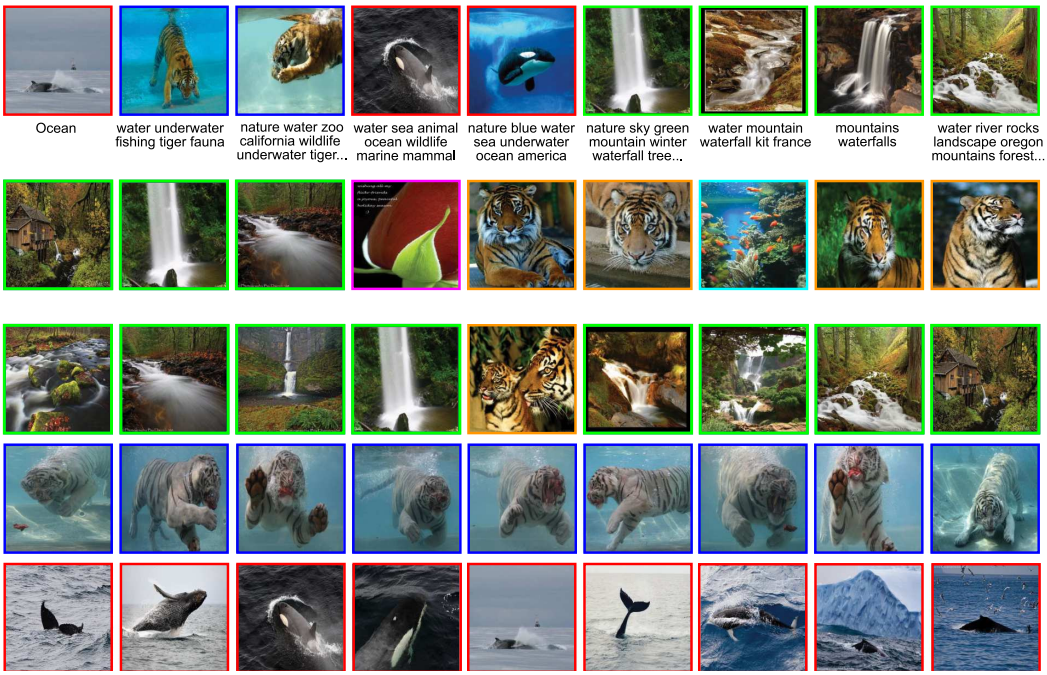


Figure 5: Spectral clustering of NUS dataset. Shown are a few images and corresponding tags belonging to the same cluster obtained using Tags (top row), Color histogram (second row), and joint modalities (third to fifth row). Groundtruth clusters are shown in different colors.

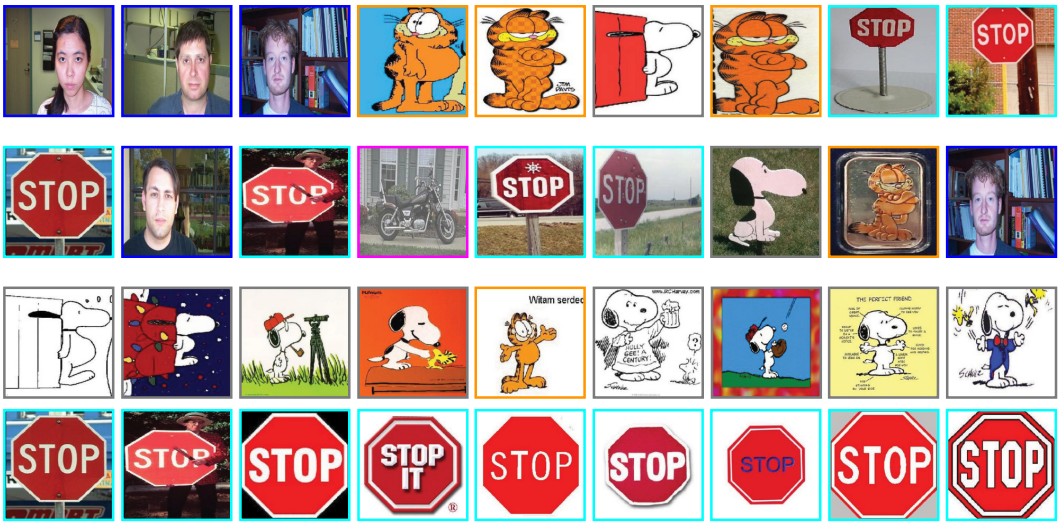


Figure 6: Spectral clustering of Caltech101 dataset. Shown are a few images and corresponding tags belonging to the same cluster obtained using ht_bio_105034 bio-inspired features (top row), 4x4 PHOW (second row), and joint modalities (third and fourth row). Groundtruth clusters are shown in different colors.

maps. In follow-up studies, we intend to show multi-modal extensions of other related techniques such as spectral hashing.

We also showed that many previously proposed approaches to multi-modal spectral clustering are nearly equivalent and try to solve some version of the joint approximate diagonalization problem. From the numerical perspective, existing methods were tailored for computing the null joint eigenvectors that are sought for in clustering problems. The underlying optimization problems are poorly suited for broader applications of diffusion geometry such as non-linear dimensionality reduction

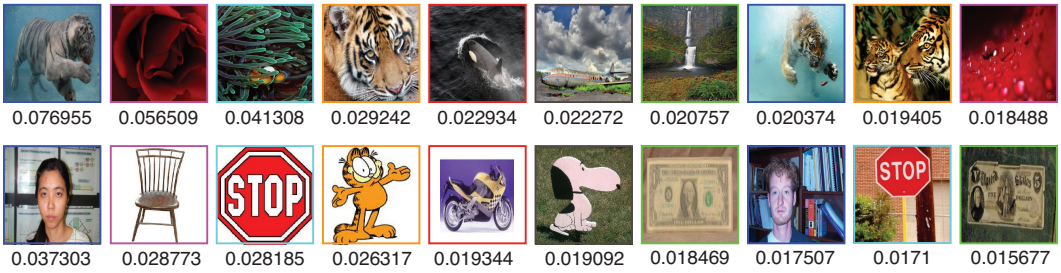


Figure 7: Farthest point sampling of NUS (top) and Caltech (bottom) datasets using joint diffusion distance. First point is on the left. Numbers indicate the sampling radius. Note that in both cases, the first 7 samples cover all the image classes.

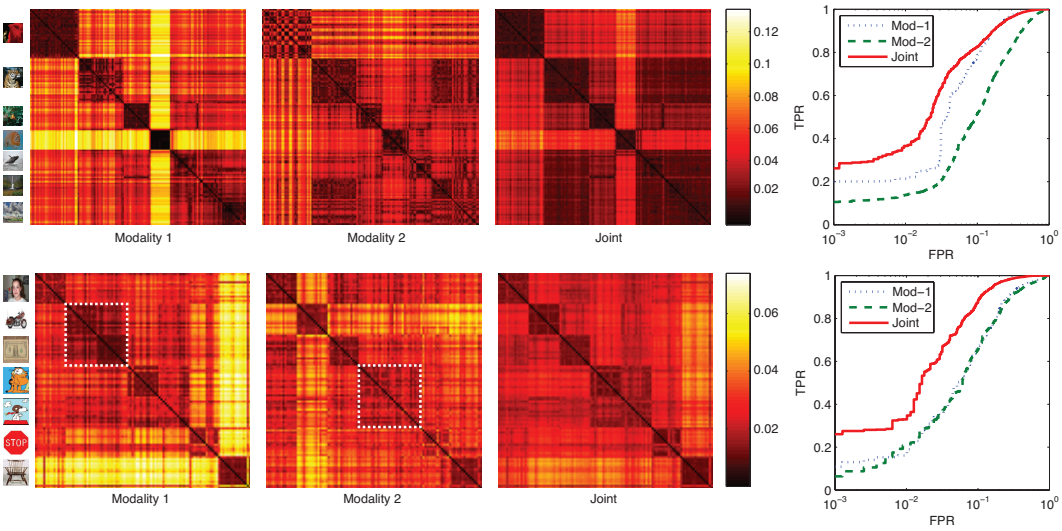


Figure 8: Columns one to three: distance matrices, column four: ROC curves, computed on NUS (top) and Caltech (bottom) datasets using joint diffusion distance. Ambiguities are shown in white.

and manifold learning, where many or all eigenvectors of the Laplacians are of interest. While approximate joint diagonalization methods developed in the signal processing community for source separation problems can address the latter case, they were initially developed for full matrices and do not take advantage of the sparse structure of Laplacians.

To the best of our knowledge, there currently exists no efficient tool to compute the joint eigenvectors of very large sparse matrices, akin Matlab's `eigs`. We believe that the presented construction makes the need of such a tool central enough to deserve the interest of the entire machine learning community. In future work, we will consider extending standard methods for eigendecomposition of large sparse matrices to the joint diagonalization case.

References

Alameda-Pineda, X., Khalidov, V., Horaud, R., and Forbes, F. Finding audio-visual events in informal social gatherings. In *Proc. ICMI*, 2011.

Bekkerman, R. and Jeon, J. Multi-modal clustering for multimedia collections. In *Proc. CVPR*, 2007.

Belkin, M. and Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15:1373–1396, 2002.

Bronstein, M. M., Bronstein, A. M., Michel, F., and Paragios, N. Data fusion through cross-modality metric learning using similarity-sensitive hashing. In *Proc. CVPR*, pp. 3594–3601, 2010.

Bunse-Gerstner, A., Byers, R., and Mehrmann, V. Numerical methods for simultaneous diagonalization. *SIAM J. Matrix Anal. Appl.*, 14(4):927–949, 1993.

Cai, X., Nie, F., Huang, H., and Kamangar, F. Heterogeneous image feature integration via multi-modal spectral clustering. In *Proc. CVPR*, 2011.

Cardoso, J.-F. and Souloumiac, A. Blind beamforming for non-gaussian signals. *Radar and Signal Processing*, 140(6):362–370, 1993.

Cardoso, J.-F. and Souloumiac, A. Jacobi angles for simultaneous diagonalization. *SIAM J. Matrix Anal. Appl*, 17:161–164, 1996.

Chua, T.-S., Tang, J., Hong, R., Li, H., Luo, Z., and Zheng, Y.-T. Nus-wide: A real-world web image database from national university of singapore. In *Proc. CIVR*, 2009.

Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Warner, F., and Zucker, S. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. In *PNAS*, pp. 7426–7431, 2005.

Coifman, R.R. and Lafon, S. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21:5–30, 2006.

de Sa, V.R. Spectral clustering with two views. In *Proc. ICML Workshop on learning with multiple views*, 2005.

Ding, C.H.Q., He, Xiaofeng, Zha, Hongyuan, Gu, Ming, and Simon, H.D. A min-max cut algorithm for graph partitioning and data clustering. In *Proc. Conf. Data Mining*, 2001.

Hochbaum, D. S. and Shmoys, D. B. A best possible heuristic for the k-center problem. *Mathematics of operations research*, pp. 180–184, 1985.

Kidron, E., Schechner, Y. Y., and Elad, M. Pixels that sound. In *Proc. CVPR*, 2005.

Kumar, A., Rai, P., and Daumé III, H. Co-regularized multi-view spectral clustering. In *Proc. NIPS*, 2011.

Levy, B. Laplace-Beltrami eigenfunctions towards an algorithm that "understands" geometry. In *Proc. SMI*, 2006.

Ma, C. and Lee, C.-H. Unsupervised anchor shot detection using multi-modal spectral clustering. In *Proc. ICASSP*, 2008.

McFee, B. and Lanckriet, G. R. G. Learning multi-modal similarity. *JMLR*, 12:491–523, 2011.

Ng, A. Y., Jordan, M. I., and Weiss, Y. On spectral clustering: Analysis and an algorithm. In *Proc. NIPS*, 2001.

Perona, P. and Zelnik-Manor, L. Self-tuning spectral clustering. In *Proc. NIPS*, 2004.

Rahbar, K. and Reilly, J. P. Geometric optimization methods for blind source separation of signals. In *Proc. ICA*, pp. 375–380, 2000.

Rasiwasia, N., Costa Pereira, J., Coviello, E., Doyle, G., Lanckriet, G.R.G., Levy, R., and Vasconcelos, N. A new approach to cross-modal multimedia retrieval. In *Proc. ICM*, pp. 251–260, 2010.

Shi, J. and Malik, J. Normalized cuts and image segmentation. *Trans. PAMI*, 22:888–905, 1997.

Tang, W., Lu, Z., and Dhillon, I.S. Clustering with multiple graphs. In *Proc. Data Mining*, 2009.

von Luxburg, U. A tutorial on spectral clustering. 2007.

Weiss, Y., Torralba, A., and Fergus, R. Spectral hashing. In *Proc. NIPS*, 2008.

Weston, J., Bengio, S., and Usunier, N. Large scale image annotation: learning to rank with joint word-image embeddings. *Machine learning*, 81(1):21–35, 2010.

Yeredor, A. Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. *Trans. Signal Proc.*, 50(7):1545 –1553, 2002.

Ziehe, A. *Blind Source Separation based on Joint Diagonalization of Matrices with Applications in Biomedical Signal Processing*. Dissertation, University of Potsdam, 2005.